数理解析研究所講究録 1224 巻 2001 年 187-193

Bayesian estimation of predictive densities

筑波大・数学 赤平 昌文 (Masafumi Akahira)
筑波大・VBL 西平 祐治 (Yuji Nishihira)
筑波大・数学 飛田 英祐 (Eisuke Hida)

1. はじめに

統計的推測理論において, 観測可能なデータに基づいて未観測な確率変数に関する何らかの知見を得る問題を統計的予測問題という. たとえば, 通常の推定論において母数の推定量に対応するものとして予測量があり, 何らかの意味で最良な予測量を求めることができることもある ([Gu70], [T75], [A90]).

最近, Bayes 的観点から未観測な確率変数の分布, すなわち予測分布についての研究が盛んに行われている. その場合に損失関数として情報量を用いることが多く, 適当な事前分布に関する Bayes リスクを最小にする予測分布を求めることができる. 本論では, まず Corcuera and Giummolè[CG99] に従って, 彼らの結果を述べ, 次に, Akahira[A96] が非正則推定論において導入した情報量を損失関数として, その Bayes リスクの下界について考察する.

2. 設定

まず, $X_1, \ldots, X_n, Y$ を互いに独立に, いずれも (σ-有限測度 μ に関する) 密度 $p_0(x;\theta)$ $(\theta \in \Theta)$ に従う確率変数とする. ただし, Θ は母数空間で, Θ は R^m の開集合とする. このとき, $(X_1, \ldots, X_n)$ の同時密度は $p(\boldsymbol{x};\theta) = \prod_{i=1}^n p_0(x_i;\theta)$ になる. ただし, $\boldsymbol{x} = (x_1, \ldots, x_n)$ とする. いま $X_1, \ldots, X_n$ を観測可能なデータ, Y を未観測な確率変数と見なし, $\boldsymbol{X} = (X_1, \ldots, X_n)$ に基づいて Y の密度 p_0 を密度

$$\hat{p}_0(y;\boldsymbol{x})$$

によって予測する. この $\hat{p}_0$ を Y の予測密度 (predictive density) という. そのような予測密度を求める通常の方法は, θ の推定量 $\hat{\theta} = \hat{\theta}(\boldsymbol{X})$ を用いて

$$\hat{p}_0(y;\boldsymbol{x}) = p_0(y;\hat{\theta}(\boldsymbol{x}))$$

とすることである. もう 1 つの方法に Bayes 法がある. まず, 母数空間 Θ 上の (ルベーグ測度に関する) 事前密度 π を与えたとき, Bayes 予測密度を

$$\hat{p}_0(y;\boldsymbol{x}) = p_0(y|\boldsymbol{x}) = \int_\Theta p_0(y;\theta)\pi(\theta|\boldsymbol{x})d\theta \tag{2.1}$$

で定義する. ただし, $\pi(\theta|\boldsymbol{x})$ は $\boldsymbol{X} = \boldsymbol{x}$ を与えたときの θ の事後密度, すなわち

$$\pi(\theta|\boldsymbol{x}) = \frac{p(\boldsymbol{x};\theta)\pi(\theta)}{\int_\Theta p(\boldsymbol{x};\theta)\pi(\theta)d\theta}$$

とする. いま, 情報量 $D(p,\hat{p})$ を損失関数として, その平均損失

$$E[D(p_0,\hat{p}_0)] = \int_{\mathcal{X}^n} D(p_0(y;\theta),\hat{p}_0(y;\boldsymbol{x}))p(\boldsymbol{x};\theta)d\mu^n(\boldsymbol{x}) \tag{2.2}$$

によって, 予測密度 $\hat{p}_0$ の良さについて考える. ただし, $\mathcal{X}$ は各 X_i の標本空間で, $\mathcal{X}^n$ は $\mathcal{X}$ の n 個の直積空間とし, μ^n を μ の n 個の直積測度とする. しかし, (2.2) は未知の p に依存するので, 事前密度 π によって (2.2) を積分すれば p に無関係にできる. よって, $\hat{p}_0$ の Bayes リスク

$$R_D(\hat{p}_0) := \int_\Theta E[D(p_0,\hat{p}_0)]\pi(\theta)d\theta \tag{2.3}$$

は, 真の密度 p との近さの尺度と見なせるであろう.

Csiszár[C67] は, $|\alpha| \leq 1$ について, α-情報量

$$D_\alpha(p_0,\hat{p}_0) := \int_{\mathcal{X}} f_\alpha\left(\frac{\hat{p}_0(y;\boldsymbol{x})}{p_0(y;\theta)}\right)p_0(y;\theta)d\mu(y),$$

を導入した. ただし,

$$f_\alpha(z) = \begin{cases} \frac{4}{1-\alpha^2}(1-z^{(1+\alpha)/2}) & (|\alpha|<1), \\ z\log z & (\alpha=1), \\ -\log z & (\alpha=-1) \end{cases}$$

とする. ここで, $\alpha=-1$ のときに D_{-1} は Kullback-Leibler 情報量, $\alpha=0$ のときに D_0 は Hellinger 距離の 2 倍の 2 乗になる. この α-情報量を損失関数として Bayes リスクを最小にする予測密度はすでに得られている ([CG99]).

一方, Akahira[A96] は非正則分布族において, 統計量の情報損失を論じた際に, 一般情報量として

$$I^{(\alpha)}(p_0,\hat{p}_0) = -\frac{8}{1-\alpha^2}\log\int_{\mathcal{X}}\left(p_0(y;\theta)\right)^{(1-\alpha)/2}\left(\hat{p}_0(y;\boldsymbol{x})\right)^{(1+\alpha)/2}d\mu(y)$$

を導入した. ただし, $-1<\alpha<1$ とする. ここで, $\alpha=0$ とすると積分は類似度 (affinity) になっていることに注意. また, 一般情報量の極限として Fisher 情報量を得ることができる.

本論では, 損失関数として α-情報量をとった場合の [CG99] の結果を踏まえて (第 3 節), 一般情報量を用いた Bayes リスクの下界について論じ (第 4 節, [AN01]), そして特に正規分布, 一様分布の場合に具体的に考察する (第 5 節).

3. α-情報量による Bayes 予測密度

母数空間 Θ 上の事前密度 π を与えたときに, 損失関数 D として α-情報量 D_α をとって, Bayes リスク (2.3) を最小にする予測密度を見つける問題を考える.

このとき, $|\alpha| \leq 1$ について $p_0(y;\theta)$ の一般 Bayes 予測密度を

$$\hat{p}_0^{(\alpha)}(y;\boldsymbol{x}) \propto \begin{cases} \left\{\int_\Theta \left(p_0(y;\theta)\right)^{(1-\alpha)/2}\pi(\theta|\boldsymbol{x})d\theta\right\}^{2/(1-\alpha)} & (\alpha\neq 1), \\ \exp\left\{\int_\Theta \log p_0(y;\theta)\pi(\theta|\boldsymbol{x})d\theta\right\} & (\alpha=1) \end{cases} \tag{3.1}$$

で定義する. ここで, $\alpha=-1$ のとき, $\hat{p}_0^{(\alpha)}$ は (2.1) と一致することに注意.

定理 3.1([CG99]). 損失関数 D として α-情報量 D_α をとったときに, $p_0(y;\theta)$ の Bayes 推定量, すなわち Bayes リスク (2.3) を最小にする推定量は, (3.1) の一般 Bayes 予測密度 $\hat{p}_0^{(\alpha)}(y;\boldsymbol{x})$ である.
証明の概略. $|\alpha|<1$ について, α-情報量は

$$D_\alpha(p_0,\hat{p}_0^{(\alpha)})=\frac{4}{1-\alpha^2}\left[1-\int_{\mathcal{X}}\left(p_0(y;\theta)\right)^{(1-\alpha)/2}\left(\hat{p}_0^{(\alpha)}(y;\boldsymbol{x})\right)^{(1+\alpha)/2}d\mu(y)\right]$$

になる. いま $\hat{r}(y;\boldsymbol{x})$ を $\hat{p}_0^{(\alpha)}$ と異なる任意の予測密度とすれば

$$\begin{aligned}&E[D_\alpha(p_0,\hat{r})-D_\alpha(p_0,\hat{p}_0^{(\alpha)})]\\ =&\frac{4}{1-\alpha^2}\int_{\mathcal{X}^n}\int_{\mathcal{X}}\left[\left(\hat{p}_0^{(\alpha)}(y;\boldsymbol{x})\right)^{(1+\alpha)/2}-\left(\hat{r}(y;\boldsymbol{x})\right)^{(1+\alpha)/2}\right]\left(p_0(y;\theta)\right)^{(1-\alpha)/2}p(\boldsymbol{x};\theta)d\mu(y)d\mu^n(\boldsymbol{x})\end{aligned}$$

となり, この辺々を事前分布 π で積分すれば

$$\begin{aligned}&\int_\Theta E[D_\alpha(p_0,\hat{r})-D_\alpha(p_0,\hat{p}_0^{(\alpha)})]\pi(\theta)d\theta\\ =&\int_{\mathcal{X}^n}D_\alpha(\hat{p}_0^{(\alpha)},\hat{r})p(\boldsymbol{x})\left\{C_\alpha(\boldsymbol{x})\right\}^{(1-\alpha)/2}d\mu^n(\boldsymbol{x})\geq 0\end{aligned}$$

になる. ただし, $C_\alpha(\boldsymbol{x})$ は $\hat{p}_0^{(\alpha)}$ の正規化定数で

$$p(\boldsymbol{x}):=\int_\Theta p(\boldsymbol{x};\theta)\pi(\theta)d\theta$$

とする. よって, $\hat{p}_0^{(\alpha)}$, $\hat{r}$ の Bayes リスクについて

$$R_{D_\alpha}(\hat{p}_0^{(\alpha)})\leq R_{D_\alpha}(\hat{r})$$

が成り立つから, $\hat{p}_0^{(\alpha)}$ は Bayes 推定量になる. また, $\alpha=\pm1$ についても同様に証明される. □

4. 一般情報量による Bayes リスクの下界

本節では, 事前密度 π を与えたときに, 損失関数 D として一般情報量 $I^{(\alpha)}$ をとって, Bayes リスク (2.3) を考える. まず, $|\alpha|<1$ について α-情報量は

$$D_\alpha(p_0,\hat{p}_0)=\frac{4}{1-\alpha^2}\left\{1-\int_{\mathcal{X}}\left(p_0(y;\theta)\right)^{(1-\alpha)/2}\left(\hat{p}_0(y;\boldsymbol{x})\right)^{(1+\alpha)/2}d\mu(y)\right\}$$

になるから, これと一般情報量 $I^{(\alpha)}$ との関係は

$$D_\alpha(p_0,\hat{p}_0)=\frac{4}{1-\alpha^2}\left[1-\exp\left\{-\frac{1-\alpha^2}{8}I^{(\alpha)}(p_0,\hat{p}_0)\right\}\right]\tag{4.1}$$

になる. このとき, 一般情報量 $I^{(\alpha)}(p_0,\hat{p}_0)$ を損失関数としたときの Bayes リスクの下界を求めよう.
定理 4.1([AN01]). 損失関数 D として一般情報量 $I^{(\alpha)}(|\alpha|<1)$ をとったとき, $p_0(y;\theta)$ の予測密度 $\hat{p}_0(y;\boldsymbol{x})$ の Bayes リスク (2.3) の下界は

$$R_{I^{(\alpha)}}(\hat{p}_0)\geq-\frac{8}{1-\alpha^2}\log\left\{1-\frac{1-\alpha^2}{4}R_{D_\alpha}(\hat{p}_0^{(\alpha)})\right\}=:B_\alpha$$

によって与えられる．ただし，$\hat{p}_0^{(\alpha)}$ は一般 Bayes 予測密度 (3.1) で $R_{D_\alpha}(\hat{p}_0^{(\alpha)})$ は α-情報量を損失関数にしたときの $\hat{p}_0^{(\alpha)}$ の Bayes リスクとする．

証明．まず，$\hat{p}_0$ の Bayes リスクは，(4.1) より

$$\begin{aligned} R_{D_\alpha}(\hat{p}_0) &= \int_\Theta E[D_\alpha(p_0,\hat{p}_0)]\pi(\theta)d\theta \\ &= \frac{4}{1-\alpha^2}\left\{1-\int_\Theta E\left[\exp\left\{-\frac{1-\alpha^2}{8}I^{(\alpha)}(p_0,\hat{p}_0)\right\}\right]\pi(\theta)d\theta\right\} \end{aligned} \tag{4.2}$$

になるから，これを最小にする $\hat{p}_0$ を求めるには

$$S_\pi(\hat{p}_0) := \int_\Theta E\left[\exp\left\{-\frac{1-\alpha^2}{8}I^{(\alpha)}(p_0,\hat{p}_0)\right\}\right]\pi(\theta)d\theta$$

を最大にする $\hat{p}_0$ を求めればよい．そこで Jensen の不等式を 2 回用いて

$$\begin{aligned} S_\pi(\hat{p}_0) &\geq \int_\Theta \exp\left\{E\left[-\frac{1-\alpha^2}{8}I^{(\alpha)}(p_0,\hat{p}_0)\right]\right\}\pi(\theta)d\theta \\ &\geq \exp\left\{-\frac{1-\alpha^2}{8}\int_\Theta E\left[I^{(\alpha)}(p_0,\hat{p}_0)\right]\pi(\theta)d\theta\right\} \\ &= \exp\left\{-\frac{1-\alpha^2}{8}R_{I^{(\alpha)}}(\hat{p}_0)\right\} \end{aligned} \tag{4.3}$$

になる．一方，定理 3.1, (4.2) より

$$R_{D_\alpha}(\hat{p}_0^{(\alpha)}) = \min_{\hat{p}_0} R_{D_\alpha}(\hat{p}_0) = \frac{4}{1-\alpha^2}\left\{1-\max_{\hat{p}_0} S_\pi(\hat{p}_0)\right\}$$

となるから，(4.3) から

$$\begin{aligned} \exp\left\{-\frac{1-\alpha^2}{8}R_{I^{(\alpha)}}(\hat{p}_0)\right\} &\leq \max_{\hat{p}_0} S_\pi(\hat{p}_0) \\ &= 1-\frac{1-\alpha^2}{4}R_{D_\alpha}(\hat{p}_0^{(\alpha)}) \end{aligned}$$

になる．よって，(4.3) を変形すれば

$$R_{I^{(\alpha)}}(\hat{p}_0) \geq -\frac{8}{1-\alpha^2}\log\left\{1-\frac{1-\alpha^2}{4}R_{D_\alpha}(\hat{p}_0^{(\alpha)})\right\}$$

定理 4.1 で与えた Bayes リスクの下界 B_α は達成可能ではないが, Bayes リスクと下界について正規分布と一様分布の場合に考えてみよう.

例 5.1. $X_1,\ldots,X_n,Y$ をたがいに独立に, いずれも正規分布 $N(\mu,\sigma^2)$ に従う確率変数とする. ただし, $-\infty<\mu<\infty$, $\sigma>0$ とする. このとき $\theta:=(\mu,\sigma^2)$ に対する十分統計量は $\hat{\mu}:=(1/n)\sum_{i=1}^n X_i=\bar{X}$, $\hat{\sigma}^2:=(1/n)\sum_{i=1}^n(X_i-\bar{X})^2$ になる. いま, θ の事前分布として $\pi(\theta)\propto 1/\sigma$ $(0<\sigma<\infty)$ をとると, 事後分布は

$$\begin{aligned}\pi(\theta|\boldsymbol{x}) &\propto p(\boldsymbol{x};\theta)\pi(\theta)\\ &\propto \left(\frac{1}{\sigma}\right)^{n+1}\exp\left\{-\frac{1}{2}\sum_{i=1}^n\left(\frac{x_i-\mu}{\sigma}\right)^2\right\}\end{aligned}$$

になる. このとき, $\alpha\neq 1$ に対して, 一般 Bayes 予測密度は

$$\hat{p}_0^{(\alpha)}(y;\boldsymbol{x})\propto\left\{\int_0^\infty\int_{-\infty}^\infty\big(p_0(y;\theta)\big)^{(1-\alpha)/2}\pi(\theta|\boldsymbol{x})d\mu d\sigma\right\}^{2/(1-\alpha)}$$

になり

$$\hat{p}_0^{(\alpha)}(y;\boldsymbol{x})=\frac{K_{n,\alpha}}{\hat{\sigma}}\left\{1+\frac{1-\alpha}{2n+1-\alpha}\left(\frac{y-\hat{\mu}}{\hat{\sigma}}\right)^2\right\}^{-\frac{2n-1-\alpha}{2(1-\alpha)}}$$

になる ([CG99], 図 5.1 参照). ただし,

$$K_{n,\alpha}=\frac{\Gamma\left(\frac{2n-1-\alpha}{2(1-\alpha)}\right)}{\sqrt{\pi}\Gamma\left(\frac{n-1}{1-\alpha}\right)}\sqrt{\frac{1-\alpha}{2n-1-\alpha}}$$

とする. この $\hat{p}_0^{(\alpha)}$ について

$$B_\alpha=-\frac{8}{1-\alpha^2}\log\left\{1-\frac{1-\alpha^2}{4}R_{D_\alpha}(\hat{p}_0^{(\alpha)})\right\}$$

となり

$$R_{I^{(\alpha)}}(\hat{p}_0^{(\alpha)})=\int_0^\infty\int_{-\infty}^\infty E[I^{(\alpha)}(p_0,\hat{p}_0^{(\alpha)})]\frac{1}{\sigma}d\mu d\sigma$$

になる. ただし

$$I^{(\alpha)}(p_0,\hat{p}_0^{(\alpha)})=-\frac{8}{1-\alpha^2}\log\int_{-\infty}^\infty\left\{\frac{1}{\sqrt{2\pi}\sigma}e^{-\frac{(y-\mu)^2}{2\sigma^2}}\right\}^{(1-\alpha)/2}\left\{\hat{p}_0^{(\alpha)}(y;\boldsymbol{x})\right\}^{(1+\alpha)/2}dy$$

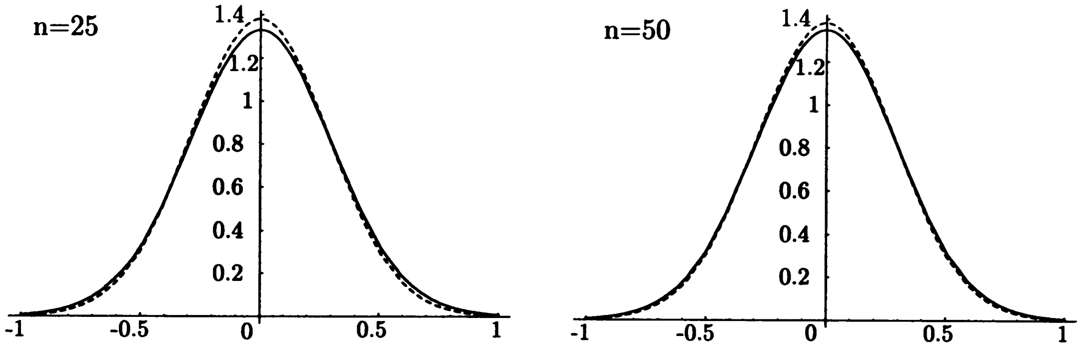

図 5.1 正規分布 $N(0,1/12)$ の密度 p_0 に対する予測密度 $\hat{p}_0^{(0)}(\cdot;\boldsymbol{x})$. ただし, $n=25$, 50 で反復数 500. ……………… $p_0(\cdot;1/2)$, ———— $\hat{p}_0^{(0)}(\cdot;\boldsymbol{x})$

例 5.2. $X_1,\dots,X_n,Y$ をたがいに独立で, いずれも一様分布 $U(\theta-(1/2),\ \theta+(1/2))$ の密度 $p_0(\cdot;\theta)$ をもつ確率変数とする. ただし, $-\infty<\theta<\infty$ とする. このとき, $\boldsymbol{X}:=(X_1,\dots,X_n)$ の同時密度は

$$p(\boldsymbol{x};\theta)=\prod_{i=1}^{n}p_0(x_i;\theta)=\chi_{[\underline{\theta},\ \bar{\theta}]}(\theta) \tag{5.1}$$

になる. ただし, $\underline{\theta}:=\max_{1\le i\le n}x_i-(1/2)$, $\bar{\theta}:=\max_{1\le i\le n}x_i+(1/2)$ とする. いま, θ の事前分布として一様分布 $U(-c,c)$ $(c>0)$ をとり, その密度を π とすれば事後密度は, (5.1) より

$$\begin{aligned}\pi(\theta|\boldsymbol{x})&\propto p(\boldsymbol{x};\theta)\pi(\theta)\\&\propto\begin{cases}\chi_{[A,B]}(\theta) & (A<B),\\ 0 & (A\ge B)\end{cases}\end{aligned} \tag{5.2}$$

になる. ただし, $\boldsymbol{x}=(x_1,\dots,x_n)$, $A:=\max\{\underline{\theta},-c\}$, $B:=\min\{\bar{\theta},c\}$ とする. このとき, $\alpha\ne 1$ に対して, 一般 Bayes 予測密度は

$$\hat{p}_0^{(\alpha)}(y;\boldsymbol{x})\propto\left\{\int_{-\infty}^{\infty}(p_0(y;\theta))^{(1-\alpha)/2}\pi(\theta|\boldsymbol{x})d\theta\right\}^{2/(1-\alpha)}$$

より, (5.2) から $A<B$ について

$$\begin{aligned}\hat{p}_0^{(\alpha)}(y;\boldsymbol{x})&\propto\left(\int_{y-\frac{1}{2}}^{y+\frac{1}{2}}\chi_{[A,B]}(\theta)d\theta\right)^{2/(1-\alpha)}\\&=\begin{cases}0 & (y\le A-\frac{1}{2}\ \text{または}\ y>B+\frac{1}{2}),\\ \left(y-A+\frac{1}{2}\right)^{2/(1-\alpha)} & (A-\frac{1}{2}<y\le\min\{A+\frac{1}{2},B-\frac{1}{2}\}),\\ 1 & (A+\frac{1}{2}<y\le B-\frac{1}{2}),\\ \left(-y+B+\frac{1}{2}\right)^{2/(1-\alpha)} & (\max\{A+\frac{1}{2},B-\frac{1}{2}\}<y\le B+\frac{1}{2}),\end{cases}\end{aligned}$$

になり, $A\ge B$ について $\hat{p}_0^{(\alpha)}(y;\boldsymbol{x})=0$ になる (図 5.2 参照). この $\hat{p}_0^{(\alpha)}$ について

$$B_\alpha=-\frac{8}{1-\alpha^2}\log\left\{1-\frac{1-\alpha^2}{4}R_{D_\alpha}(\hat{p}_0^{(\alpha)})\right\}$$

$$R_{I^{(\alpha)}}(\hat{p}_0^{(\alpha)}) = \frac{1}{2c}\int_{-c}^{c} E[I^{(\alpha)}(p_0,\ \hat{p}_0^{(\alpha)})]d\theta$$

になる．ただし

$$I^{(\alpha)}(p_0,\ \hat{p}_0^{(\alpha)}) = -\frac{8}{1-\alpha^2}\log\int_{\theta-\frac{1}{2}}^{\theta+\frac{1}{2}}\{\hat{p}_0(y;\boldsymbol{x})\}^{(1+\alpha)/2}dy$$

とする．

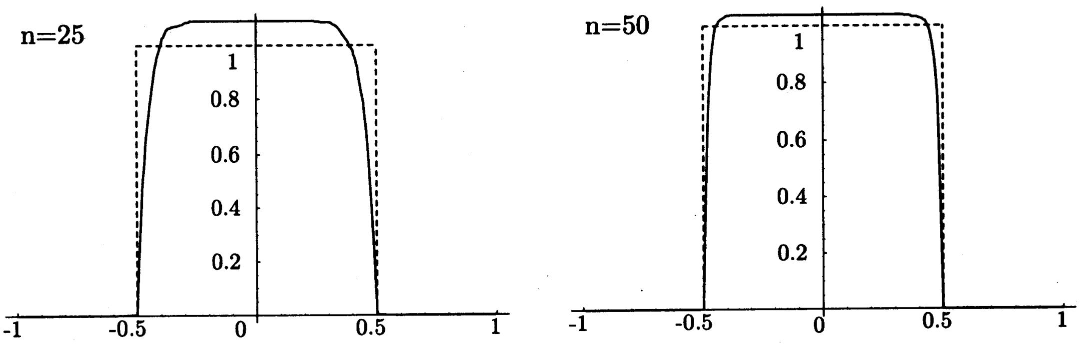

図 5.2 一様分布 $U(-1/2, 1/2)$ の密度 $p_0(\cdot;1/2)$ に対する予測密度 $\hat{p}_0^{(0)}(\cdot;\boldsymbol{x})$．ただし，$n=25, 50$ で反復数 500.　……………… $p_0(\cdot;1/2)$,　———— $\hat{p}_0^{(0)}(\cdot;\boldsymbol{x})$

参考文献

[A90] Akahira, M. (1990). *Theory of Statistical Prediction.* Lecture Note at the Middle East Technical University, Ankara.

[A96] Akahira, M. (1996). Loss of information of a statistic for a family of non-regular distributions. *Ann. Inst. Statist. Math.*, **48**, 349–364.

[AN01] Akahira, M. and Nishihira, Y. (2001). The lower bound for the Bayes risk of predictive densities. In preparation.

[CG99] Corcuera, J. M. and Giummolè, F. (1999). A generalized Bayes rule for prediction. *Scand. J. Statist.*, **26**, 265–279.

[Cs67] Csiszár, I. (1967). Information-type measures of difference of probability distributions and indirect observations. *Studia Sci. Math. Hung.*, **2**, 299–318.

[G93] Geisser, S. (1993). *Predictive Inference : An Introduction.* Chapman & Hall, New York.

[Gu70] Guttman, I. (1970). *Statistical Tolerance Regions : Classical and Bayesian.* Griffin, London.

[T75] 竹内啓. (1975). 統計的予測論. 培風館.